

ASA-927  
E5397-01EJ

Title of the Invention

METHOD AND SYSTEM FOR RETRIEVING A DOCUMENT AND  
COMPUTER READABLE STORAGE MEDIUM

Inventors

Yasuhiko INABA  
Katsumi TADA  
Natsuko SUGAYA  
Tadataka MATSUBAYASHI  
Akihiko YAMAGUCHI  
Mikihiko TOKUNAGA

METHOD AND SYSTEM FOR RETRIEVING A DOCUMENT AND  
COMPUTER READABLE STORAGE MEDIUM

## BACKGROUND OF THE INVENTION

The present invention relates to method and system for retrieving a document on the basis of a search condition designated by a user and a storage medium storing a processing program for use in the method and system and more particularly, to a retrieval technique in which the user evaluates a document obtained as a result of retrieval and the evaluation is reflected on the search condition.

10 With the widespread use of personal computers, Internet and the like, the number of electronic documents has been increasing drastically. Under the circumstances, a demand for retrieving a document containing information desired by a user at a high speed and with high efficiency has been increasing.

There is available a technique called relevance feedback as a retrieval technique meeting the demand as above. In this technique, for a search result based on a whole passage search or similar document search, the user inputs an evaluation "the document is desirable" or "the document is undesirable" to the system so that the evaluation information may be reflected to modify the search condition so as to improve the succeeding search results.

25 Specifically, a process method is shown in,

for example, "Information Retrieval" by William B. Frakes/Rocardo Baeza-Yates, Prentice Hall PTR, 1992 pp.241-263, according to which weights in a search condition for words extracted from a document that has been evaluated as desirable by the user add and weights in a search condition for words extracted from a document that has been evaluated as undesirable subtract. Hereinafter, the above technique will be referred to as related art 1.

10           An example of a concrete method of addition/subtraction of weight applied to a certain word in a search condition is given by equation (1).

$$W' = W + \alpha \sum_i^P Fp(i) - \beta \sum_j^N Fn(j) \quad \dots\dots (1)$$

where  $W'$  represents a weight of the word after feedback has been carried out,  $W$  represents a weight of the word before the execution of feedback,  $Fp(i)$  represents the appearance frequency (times) of the word in an  $i$ -th document evaluated as desirable, and  $Fn(j)$  represents the appearance frequency of the word in a  $j$ -th document evaluated as undesirable. In the equation,  $P$  is the number of documents evaluated as desirable,  $N$  is the number of documents evaluated as undesirable, and  $\alpha$  and  $\beta$  are parameters. The new weight  $W'$  can have a negative value and in such a case, similarity of the document containing the word decreases.

In the related art 1, however, there arises a problem to be described below. More particularly, the related art 1 employs a scheme in which one search condition changes with evaluations by the user, but the  
5 related art 1 does not disclose a means for saving a history of changes and a history of user's evaluations by making the correspondence between them. Consequently, once an evaluation made by the user is erased, it is impossible to return the search condition to that at a  
10 certain time.

In the relevance feedback, feedback is carried out by selecting a document to be evaluated from restricted documents contained in a set of search results and there is a possibility that the feedback  
15 does not always meet an intention of the user. In such an event, the search result is rather degraded by the feedback and it is necessary to erase an evaluation so as to recommence the relevance feedback and approach an intended search result through trial and error.  
20 Disadvantageously, it is essentially impossible for the related art 1 to meet the requirements as above.

This will be exemplified by making reference to Fig. 2. The figure shows an instance where a user, who retrieves a document concerning "high-school  
25 baseball", specifies a sample document "A high-school baseball opens following soccer---" and performs a similar document search. At that time, strings such as "high-school", "baseball", "soccer" and "opens" are

extracted from the sample document and applied with weights to generate a search condition. Then, on the basis of the search condition, scores of individual documents in a database are calculated and a similar  
5 document that keeps scores satisfying a predetermined condition is delivered as a similar document search result.

An instance will now be considered in which of documents delivered as search result documents, a  
10 document "A professional baseball and a high-school long-distance relay race ---" is evaluated as "undesirable" by the user. At that time, strings such as "professional", "baseball", "high-school" and "long-distance relay race" are extracted from the document  
15 evaluated as "undesirable" and weights of these strings in the search condition subtract, with the result that a document concerning "high-school baseball" cannot thereafter be obtained.

In such an event, the related art 1 does not  
20 save a search condition before the document "A professional baseball and a high-school long-distance relay race---" is evaluated as undesirable and consequently, a search condition cannot be returned to that before changing (that is, the search condition  
25 before changing cannot be restored) by erasing the evaluation. Accordingly, for example, there arises a problem that when the relevance feedback has been repeated and then an impertinent evaluation is

inadvertently inputted, the erroneous evaluation cannot be corrected unless re-designating the initial sample document is recommenced by clearing all the search conditions.

5           Further, once the search result is degraded owing to changing of search condition, the search condition before the feedback cannot be restored and as a result, a desired document cannot be obtained as a search result. Consequently, a failure to consult the  
10 contents arises and in addition, the relevance feedback is carried out such that the document is evaluated as "desirable", giving rise to a problem that the search result cannot be improved purposively.

#### SUMMARY OF THE INVENTION

15           An object of the present invention is to realize improvements in the retrieval accuracy by making it possible to recommence evaluation for a desired search result document.

          According to the present invention, in a  
20 document retrieval method in which a user inputs an evaluation for evaluating a search result document based on a designated search condition as desirable or undesirable and the search condition is changed on the basis of the evaluation, when changing the search  
25 condition on the basis of the evaluation, a search condition available immediately before or after the change is saved. A search result based on the search

condition preceding or succeeding the change is saved. Further, some or all of information saved in the search condition saving step, information stored in the search result saving step, the above evaluation and an  
5 identifier of a document subjected to the evaluation are saved by making the correspondence between them.

With this construction, in a document retrieval method having the relevance feedback function, an evaluation for a desired search result document can  
10 be recommenced, thus realizing improvements in the accuracy of retrieval effected through the trial and error relevance feedback.

To accomplish the above object, there is provided a program for implementing the aforementioned  
15 function or a recording medium storing the program.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram showing the construction of an embodiment of the invention.

Fig. 2 is a diagram showing an example of a  
20 relevance feedback process representing a related art.

Fig. 3 is a PAD showing a processing procedure in a system control program 110 in the embodiment of the invention.

Fig. 4 is a PAD showing a processing procedure  
25 in a document-registering program 112 in the embodiment of the invention.

Fig. 5 is a PAD showing a processing procedure

in a search control program 113 in the embodiment of the invention.

Fig. 6 is a PAD showing a processing procedure in a searching profile generating program 114 in the  
5 embodiment of the invention.

Fig. 7 is a PAD showing a processing procedure in a similar document-searching program 115 in the embodiment of the invention.

Fig. 8 is a PAD showing a processing procedure  
10 in a search result document content displaying program 119 in the embodiment of the invention.

Fig. 9 is a PAD showing a processing procedure in a searching profile updating program 117 in the embodiment of the invention.

Fig. 10 is a PAD showing a processing  
15 procedure in a search history saving program 118 in the embodiment of the invention.

Fig. 11 is a PAD showing a processing  
procedure in a search history-displaying program 121 in  
20 the embodiment of the invention.

Fig. 12 is a PAD showing a processing  
procedure in a search result history-displaying program  
120 in the embodiment of the invention.

Fig. 13 is a PAD showing a processing  
25 procedure in a searching profile restoring program 122 in the embodiment of the invention.

Fig. 14 is a PAD showing a processing  
procedure in a user operation-erasing program 123 in the



embodiment of the invention.

Fig. 15 is a PAD showing a processing procedure in a search session ending program 130 in the embodiment of the invention.

5           Fig. 16 is a diagram showing a concrete process flow in the case where when the user initially inputs a seed document as search condition, a searching profile is generated to conduct a similar document search in the embodiment of the invention.

10           Fig. 17 is a diagram showing a process flow in the case where the user inputs an evaluation for a search result document and a searching profile is updated on the basis of the evaluation to perform re-retrieval in the embodiment of the invention.

15           Fig. 18 is a diagram showing a process flow in the case where after the user has conducted the relevance feedback to change the searching profile, the content of a searching profile at a desired time is restored and a search result based on the restored  
20   searching profile is displayed to the user in the embodiment of the invention.

Fig. 19 is a diagram showing a process flow continuing from the process flow in Fig. 18.

Fig. 20 is a diagram showing an example of a  
25   picture displayed when the user makes a request for search history display in the embodiment of the invention.

## DESCRIPTION OF THE EMBODIMENTS

An embodiment of the present invention will now be described.

In the present embodiment, retrieval is  
5 carried out on the basis of data (hereinafter referred to as searching profile) holding elements indicative of the contents a user desires in a similar document retrieval system having the relevance feedback function. A history of evaluations for search result documents  
10 inputted by the user, a history of searching profiles changed with the evaluations and a history of search results based on the changed searching profiles are all managed by making the correspondence of them with each other. Through this, a desired evaluation designated by  
15 the user can be recommenced so that the searching profile can be returned to a searching profile at a desired time. As a result, a search result document based on the returned searching profile can be presented and an evaluation for the search result document can be  
20 recommenced. In other words, improvements in accuracy of retrieval based on trail and error relevance feedback can be realized.

The present embodiment of the invention will now be described in greater detail with reference to the  
25 accompanying drawings.

Referring first to Fig. 1, a system according to the present embodiment is constructed as illustrated therein.

The system of the present embodiment comprises a display 100, a keyboard 101, a central processing unit (CPU) 102, a magnetic disc unit 105, a floppy disc drive (FDD) 106, a main memory 109 and a bus 108 for coupling  
5 these components.

The magnetic disc unit 105 is one of secondary storage units and houses a registered document storage file 103, an index for retrieval 104, a searching profile history file group 128, and a search result  
10 history file group 129. The main memory 109 or the magnetic disc unit 105 reads information stored in a floppy disc 107 through the FDD 106.

The index for retrieval 104 is a file that saves information indicating what sort of character or  
15 string in each document stored in the registered document storage file 103 appears at which position in each document. During retrieval, information as to which position in what sort of document a string appears at can be extracted at a high speed by making reference  
20 to this index.

As will be described later, the searching profile history file group 128 is a set of files each saving a searching profile each time that a searching profile is updated when the user inputs an evaluation  
25 for a search result.

As will be described later, the search result history file group 129 is a set of files each saving a search result presented to the user each time that

retrieval is carried out.

Stored in the main memory 109 are a system control program 110, a document registering program 112 and a search control program 113. The search control  
5 program 113 includes a searching profile generating program 114, a similar document searching program 115, a search result document content displaying program 119, a searching profile updating program 117, a search history saving program 118, a search history displaying program  
10 121, a search result history displaying program 120, a searching profile restoring program 122, a user operation erasing program 123 and a search session ending program 130.

The "user operation" termed herein is an  
15 operation conducted by the user for the purpose of modifying the searching profile, including inputting an evaluation for evaluating a search result document as either "desirable" or "undesirable" or recommencing the evaluation.

20 The main memory 109 also holds a searching profile saving area 124, a search history table 126 and a work area 131.

The searching profile saving area 124 is an area for saving data for retrieval that holds contents  
25 desired by the user as elements as described previously. In the present embodiment, some search strings representing the contents desired by the user and their weights are held as shown in Fig. 16 to be described

later.

The search history table 126 is a table for saving a user operation history by making the correspondence of a searching profile updated by the user operation with a search result based on the searching profile.

Processing procedures in the individual programs in the present embodiment will now be described.

10 Firstly, a processing procedure in the system control program 110 will be described by making reference to a PAD (problem analysis diagram) of Fig. 3.

The system control program 110 first analyzes, in step 401, a command inputted by the user through the keyboard 101.

Next, when the command analyzed in the step 401 is determined to be a command of document registration in step 402, the document registering program 112 is started in step 404 to register a document.

When the analyzed command is determined to be a command of search execution in step 403, the search control program 113 is started in step 405 to retrieve a document.

25 The processing procedure in the system control program 110 is carried out as described above.

Next, the document-registering program started by the system control program 110 in the step 404 shown

in Fig. 3 will be described by making reference to a PAD of Fig. 4.

In step 501, the document-registering program 112 first reads data of a document to be registered from the floppy disc 107 inserted in the FDD 106 and saves the readout document data in the registered document storage file 103 while assigning a document number for definitely identifying each document to the readout data. In an alternative, document data described as being inputted by using the floppy disc 107 may be inputted in a different way using, for example, a communication line or a CD-ROM unit (not shown in Fig. 1).

Subsequently, in step 502, index for retrieval 104 serving as data for high-speed extraction of how many times a search string in the searching profile saving area 124 appears in what sort of document is generated in respect of individual documents to be registered.

The index for retrieval 104 holds information as to which position of a registered document a character or a string appearing in the registered document appears at. For generation of the index for retrieval 104, a method as disclosed in "JP-A-64-35627" for simply extracting a string of n successive characters (n being a predetermined arbitrary value) or a different method using learning data based on neural network might be used. During retrieval, information

indicating which position of what sort of document a string appears at can be extracted at a high speed by consulting this index.

The processing procedure in the document-  
5 registering program 112 is carried out as described above.

Next, a processing procedure in the search control program 113 started by the system control program 110 in the step 405 shown in Fig. 3 will be  
10 described by using a PAD of Fig. 5.

This program is to control a relevance feedback process, in which a searching profile is modified on the basis of information inputted by the user and retrieval is repeated until the user requests a  
15 search session to be ended.

Firstly, in step 601, the search control program 113 starts the searching profile generating program 114 to generate a searching profile on the basis of a sample document (hereinafter referred to as seed  
20 document) inputted as a search condition by the user and expressing the contents desired by the user.

Next, in step 681, the similar document-searching program 115 is started to perform a document search on the basis of the searching profile generated  
25 in the step 601 and display a search result.

Thereafter, in step 682, the search history saving program 118 is started by inputting user operation information "the seed document is set", and

the searching profile and the search result are saved in a file of the searching profile saving file group 128 and a file of the search result history saving file group 129, respectively. The user operation information referred to herein is a string expressing the content of a user operation such as "the seed document is set" or "a document of document number 10 is evaluated as desirable". Through the present step, the searching profile at present and the search result based on the searching profile are saved. Further, the searching profile at present, the search result based on the searching profile and the user operation information "the seed document is set" are all registered as initial information in the search history table 126 by making the correspondence between them. The concrete contents of data saved in the search history table 126 will be described later with reference to Fig. 16.

Next, in step 602, processes in steps 603 through 618 are repeated until the search session ending program 130 is executed in step the step 618.

In the repetitive process, a command inputted from the keyboard 101 is first analyzed in the step 603.

Next, in the step 605, when the command is determined to be a command to display the content of the search result document, the search result document content displaying program 119 is started in the step 611 to display the content of a designated search result document is displayed.



Subsequently, when the command is determined to be a command to input a user's evaluation for the search result document in the step 606, the searching profile updating program 116 is started in step 651 to  
5 generate user operation information on the basis of the evaluation inputted by the user and change the content of the searching profile saving area 124. Further, in step 652, the similar document-searching program 115 is started to perform document search on the basis of the  
10 content of the searching profile changed in the step 651 and display a search result. Furthermore, in step 653, the search history saving program 118 is started by using the user operation information generated in the step 651 as an input to save, in the search history  
15 table 126, the searching profile changed in the step 651 and the search result obtained in the step 652 by making the correspondence of them with the user operation information.

Through the steps 651 to 653, the searching  
20 profile is updated on the basis of the user's evaluation for the search result document and the evaluation, the updated searching profile and the search result based on the updated searching profile are saved in the search history table 126 by making the correspondence between  
25 them.

Subsequently, when the command is determined to be a command to display a search history in the step 608, the search history displaying program 121 is

started in the step 614 to display a history of user operations following the initial seed document setting in a tabulated form.

Thereafter, in the step 607, when the command  
5 is determined to be a command to display a search result history, the search result history displaying program 120 is started in the step 613 to display search results obtained by user operations following the initial seed document setting in a tabulated form.

10 Next, in the step 609, when the command is determined to be a command to restore a searching profile, the searching profile restoring program 122 is started in step 661. In the restoration of the searching profile, with the aim of recommencing the user  
15 action from a desired past time, the content of a searching profile is returned to that at a designated time. In the present step, user operation information is generated on the basis of a request for restoration inputted by the user and besides, the content of the  
20 searching profile saving area 124 at the requested time is restored and a similar document search result based on the restored searching profile is displayed.

Further, in step 663, the search history saving program 118 is started by using an input of the user operation  
25 information generated in the step 661 to save, in the search history table 126, the searching profile changed in the step 661 and the search result obtained in the step 661 by making the correspondence of them with the

user operation information.

Through the steps 661 and 663, the searching profile is updated on the basis of the request for restoring the searching profile made by the user and the request, the updated searching profile and the search result based on the searching profile are saved in the search history table 126 by making the correspondence between them.

Next, when the command is determined to be a command to erase the user operation in the step 610, the user operation-erasing program 123 is started in step 671. In the erase of user operation, the content of searching profile is changed to that at the time that only designated one of the past user operations conducted by the user might not carried out. In the present step, user operation information is generated on the basis of a request for erase inputted by the user and besides, the content of the searching profile saved in the searching profile saving area 124 is changed to that of the searching profile obtained when the designated evaluation is not carried out. Further, in step 672, the similar document search program 115 is started to carry out document search on the basis of the content of searching profile changed in the step 671 and display a search result. Further, in step 673, the search history saving program 118 is started by using an input of the user operation information generated in the step 671 to save the searching profile changed in the

step 671 and the search result obtained in the step 672 by making the correspondence of them with the user operation information.

Through the steps 671 to 673, the searching  
5 profile is updated on the basis of the request for erasing evaluation made by the user and the request, the updated searching profile and the search result based on the searching profile are saved in the search history table 126 by making the correspondence between them.

10 Next, in the step 617, when the command is determined to be a command to end a search session, the search session ending program 130 is started in the step 618 to end the search session. In this step, the searching profile and the search history are all  
15 cleared.

In this phase, the content of the searching profile saving area 124 may be saved before cleared, so that it can be reutilized in the next and ensuing search sessions.

20 The processing procedure in the search control program 113 is carried out as described above.

Next, a processing procedure in the searching profile generating program 114 started by the search control program 113 in the step 601 shown in Fig. 5 will  
25 be described by using a PAD of Fig. 6.

In case the user inputs a seed document in the initial phase, the present program generates a searching profile on the basis of the seed document.

Firstly, in step 701, the searching profile generating program 114 reads a seed document inputted from the keyboard 101 and stores the readout seed document in the work area 131.

5           Subsequently, in step 702, strings each having a possibility of being an independent word (hereinafter referred to as characteristic strings) are extracted from the seed document stored in the work area 131 and a frequency (the number of times) of appearance of each  
10   characteristic string in the seed document is calculated as a weight used for similar document search.

Thereafter, in step 703, the characteristic strings extracted in the step 702 and their weights are written in the searching profile saving area 124.

15           Information now prepared in the searching profile saving area 124 is used when similarity of individual documents stored in the registered document storage file 103 to the seed document is calculated in the similar document searching program as will be  
20   described later. While, in this phase, the frequency of appearance in the seed document is calculated as weight in the step 702, the weight may be calculated in a different way, for example, by performing such calculation that the smaller the number of documents in  
25   registered document storage file 103 in which the individual characteristic strings appear, the higher the weight becomes.

Further, in this phase, all extracted

characteristic strings may be handled as strings to be written in the searching profile saving area 124 or a limited number of extracted characteristic strings having higher frequencies of appearance in the seed document may be selected in order of their frequencies.

For the extraction of characteristic strings, a method that is disclosed in "JP-A-11-143902", a method for extracting words in each document by using morphological analysis or other methods may be employed.

The processing procedure in the searching profile generating program is carried out as described above.

Next, a processing procedure in the similar document searching program 115 started by the search control program 113 in the step 681, 652 or 672 shown in Fig. 5 will be described with reference to a PAD of Fig. 7.

In the present program, a score conforming to conceptive similarity to a designated seed document is given to individual registered documents and a document having high similarity is displayed as a search result.

Firstly, in step 801, the similar document-searching program 115 reads the characteristic strings and their weights saved in the searching profile saving area 124 by the searching profile generating program 114 in the step 702 shown in Fig. 6.

Next, in step 806, processes in steps 802 and 803 are repeated for all characteristic strings

extracted in the step 801.

In the repetitive process, the index for retrieval 104 is first consulted in the step 802 to acquire how many times the strings appear in what sort  
5 of document.

Subsequently, in the step 803, similarity of each document adds on the basis of the appearance frequency obtained in the step 802. As a calculation expression for calculating similarity adding in each  
10 document through the repetitive process, equation (2) as below, for example, is used.

$$S(D) = \sum_i^N \{Frq(i) \times W(i)\} \dots\dots (2)$$

where S(D) represents similarity of a document number D in the registered document storage file 103, Frq(i) represents the appearance frequency of a word i in the  
15 document number D and w(i) represents a weight of the word i in the searching profile saving area 124. Obviously, another similarity calculation expression may be used.

Thereafter, in step 804, document numbers of  
20 the individual documents in the registered document storage file 103 are sorted in a descending order of similarity and then delivered to the display 100. In this phase, only predetermined document numbers of higher similarity may be delivered sequentially in a

descending order from high to low or only document numbers of higher similarity than a predetermined level of similarity may be delivered. If the document has an attribute such as title, the attribute may be delivered.

5           The processing procedure in the similar document-searching program 115 is carried out as described above.

          Referring now to a PAD of Fig. 8, a processing procedure in the search result document content  
10 displaying program 119 started by the search control program 113 in the step 611 shown in Fig. 5 will be described.

          The present program is to display the contents of a document designated by the user on the display 100.

15           Firstly, in step 901, the search result document content displaying program 119 reads a document number inputted by the user through the keyboard 101 and corresponding to a document having its contents to be displayed.

20           Next, in step 902, the document corresponding to the document number inputted in the step 901 is read out of the registered document storage file 103 and displayed on the display 100. At that time, of strings in the document to be displayed, characteristic strings  
25 stored in the searching profile saving area may be displayed with emphasis in a color different from that of other strings.

          The processing procedure in the search result



document content displaying program 119 is carried out as described above.

Referring now to a PAD of Fig. 9, a processing procedure in the searching profile updating program 117  
5 started by the search control program 113 in the step 651 shown in Fig. 5 will be described.

When a user's evaluation for a search result document is inputted, the present program feeds back the evaluation to change the content of a searching profile.

10 Firstly, in step 1501, the searching profile updating program 117 reads, as an input, a document number of a document for which the user has inputted an evaluation and the content of evaluation for evaluating the document as "desirable" or "undesirable".

15 Subsequently, in step 1502, the document corresponding to the document number read in the step 1501 is read out of the registered document storage file 103 and then, characteristic strings are extracted from the document and the frequencies of their appearance in  
20 the document are calculated.

Thereafter, in step 1507, the characteristic strings extracted in the step 1502 and the frequencies of their appearance in the document are stored in the work area 131.

25 Then, when the user's evaluation read in the step 1501 is determined to be "desirable" in step 1503, the appearance frequencies of individual characteristic strings stored in the work area 131 are added to weights

of the strings in the searching profile saving area 124 in step 1505. At that time, if the strings do not exist in the searching profile saving area 124, these strings are added to the searching profile saving area 124 and  
5 the appearance frequencies read in the step 1502 are assigned as weights

Next, when the user's evaluation read in the step 1501 is determined to be "undesirable" in step 1504, the appearance frequencies of individual characteristic strings stored in the work area 131 are  
10 subtracted from weights of the strings in the searching profile saving 124 in step 1506. At that time, if the strings do not exist in the searching profile saving area 124, these strings are added to the searching  
15 profile saving area 124 and negative values of the appearance frequencies read in the step 1502 are assigned as weights.

Next, in step 1508, user operation information is prepared on the basis of the content read in the step  
20 1501 and delivered in return to the search control program 113 that has called the present program. The user operation information prepared herein has the contents "the document of document number 10 is  
evaluated as desirable", indicating that the user has  
25 inputted an evaluation for the search result document. The search history saving program 118 scheduled to be executed after completion of execution of the present program saves that information in the search history

table 126 by making the correspondence of the information with the searching profile changed in the step 1505 or 1506 and the search result based on the searching profile.

5           In an alternative, for addition/subtraction of weight in the step 1505 or 1506, a method may be employed to define some grades in accordance with the degree of evaluation by the user. For example, when the user evaluates the document as "desirable", the  
10 appearance frequencies of the characteristic strings in the document, as they are, are added to weights of the strings in the searching profile saving area 124 but when the user evaluates the document as "somewhat desirable", half of the appearance frequencies of the  
15 characteristic strings in the document are added to weights of the strings in the searching profile saving area 124. Further, in the addition/subtraction of the characteristic strings to/from the weight in the steps 1505 and 1506, all of the characteristic strings  
20 extracted in the step 1502 may be involved or a limited number of characteristic strings having higher appearance frequencies in order may be involved.

          The processing procedure in the searching profile updating program 117 is carried out as described  
25 above.

          Referring now to a PAD of Fig. 10, a processing procedure in the search history saving program 118 started by the search control program 113 in

the step 653, 663 or 673 will be described.

The present program is called after the content of a searching profile is updated by a user operation and retrieval is executed on the basis of an  
5 updated searching profile and is to save the content of the updated searching profile and a search result based on that searching profile by making the correspondence of them with the user operation.

Firstly, in step 1601, the search history  
10 saving program 118 acquires an input of user operation information. In other words, information as to what sort of operation has changed the searching profile is acquired.

Next, in step 1607, a user operation  
15 identifier for definitely identifying the user operation information acquired in the step 1601 is prepared.

Next, in step 1603, the content of the searching profile saving area 124 is saved in a file in the searching profile history saving file group 128.

20 Then, in step 1606, a list of search results displayed on the display 100 is saved in a file in the search result history file group 129.

Subsequently, in step 1604, the user operation information acquired in the step 1601, the user  
25 operation identifier prepared in the step 1607, the file used to save the searching profile in the step 1603 and the file used to save the search result list in the step 1606 are all saved in the search history table 126 by

making the correspondence between them. The concrete contents of data saved in the search history table 126, the searching profile history saving file group 128 and the search result history file group 129 will be  
5 described later by making reference to Fig. 16.

The processing procedure in the search history saving program 118 is carried out as described above.

Referring now to a PAD of Fig. 11, a processing procedure in the search history-displaying  
10 program 121 started by the search control program 113 in the step 614 shown in Fig. 5 will be described.

After initial setting of the seed document by the user, the user has carried out user operations including inputting an evaluation for a search result  
15 document, restoring a searching profile and erasing a user operation. The present program displays a history of these user operations in a tabulated form on the display 100. For example, when the user desires to erase a past operation or reacquire a search result  
20 obtained in the past, the user operation identifier used for designating what sort of operation is to be erased and a search result at what time is to be acquired is displayed together with the user operation information by means of the present program.

25 In step 1201, the retrieval history-displaying program 121 first reads a history of user operation identifier and user operation information from the search history table 126.

Next, in step 1202, the history of user operation identifier and user operation information read in the step 1201, that is, information as to which evaluation is made for what sort of document is  
5 displayed in a list form on the display 100. In this phase, if the document subject to the evaluation has an attribute such as title, then the attribute will be delivered concurrently.

The processing procedure in the search  
10 history-displaying program 121 is carried out as described above.

Referring now to a PAD of Fig. L2, a processing procedure in the search result history-displaying program 120 started by the search control  
15 program 113 in the step 613 shown in Fig. 5 will be described.

The present program is to acquire a search result immediately after completion of a designated one of user past operations.

20 Firstly, in step 1101, the search result history-displaying program 120 starts the search history-displaying program 121 to display a search history. At that time, a list of user operations executed by the user is displayed together with user  
25 operation identifiers assigned to the operation, respectively, on the display 100 as shown in Fig. 18 to be described later.

Next, in step 1102, a user operation

identifier inputted from the keyboard 101 is read. The user operation identifier inputted at that time corresponds to a user's desire for acquiring a search result immediately after or immediately succeeding a user operation corresponding to that user operation identifier.

Subsequently, in step 1103, a file name in search result history file group 129 that corresponds to the user operation identifier read in the step 1102 is read out of the search history table 126.

Thereafter, in step 1104, a list of search results written in the search result history file corresponding to the file name read in the step 1103 is read and displayed on the display 100.

The processing procedure in the search result history-displaying program 120 is carried out as described above.

Referring now to a PAD of Fig. 13, a processing procedure in the searching retrieval profile-restoring program 122 started by the search control program 113 in the step 661 shown in Fig. 5 will be described.

In the present program, the content of searching profile is replaced with the content immediately succeeding a designated user operation, so that a different user operation starting from a designated time can be performed, that is, relevance feedback can be recommenced.

Firstly, in step 1301, the searching profile-restoring program 122 reads an input of a user operation identifier subjected to restoration. For example, in case the user desires to replace the content of  
5 searching profile with the content immediately succeeding a user operation "the document of document number 10 is evaluated as desirable", a user operation identifier assigned to this user operation is inputted.

Next, in step 1303, the search history saving  
10 area 126 is consulted to read a searching profile corresponding to the user operation identifier read in the step 1301 from the searching profile history file group 129. Namely, the content of a searching profile immediately succeeding the content changed by the user  
15 operation corresponding to the user operation identifier read in the step 1301 can be obtained.

Subsequently, in step 1304, the content in the searching profile saving area 124 is replaced with the searching profile read in the step 1303.

20 Thereafter, in step 1307, user operation information is prepared on the basis of the user operation identifier obtained in the step 1301 and is returned, as an output, to the search control program 113 that has called the present program. The user  
25 operation information now prepared has the content "a searching profile immediately succeeding the user operation of user operation identifier 5 is desired to be restored", indicating that the user desires



restoration of the searching profile. This information is saved in the search history table 126 while being caused to have the correspondence with the searching profile changed in the step 1304 and the search result  
5 based on that searching profile by means of the search history saving program scheduled to be executed after execution of the present program.

The processing procedure in the searching profile-restoring program 122 is carried out as  
10 described above.

Referring now to a PAD of Fig. 14, a processing procedure in the user operation-erasing program 123 started by the search control program 113 in the step 671 shown in Fig. 5.

15 The present program is to change the content of searching profile to the content available when only a designated one of user operations executed so far is not carried out.

Firstly, in step 1400, the user operation-  
20 erasing program reads, as an input, a user operation identifier corresponding to an operation subject to erasing.

Subsequently, in step 1401, a user operation identifier corresponding to a user operation that  
25 precedes, by one, a user operation corresponding to the user operation identifier read in the step 1400 is inputted and the searching profile restoring program 122 is started to restore the content of the searching

profile saving area 124 immediately preceding the user operation corresponding to the user operation identifier read in the step 1400.

Next, in step 1402, processes in steps 1403 to  
5 1408 are repeated for user operation identifiers ranging over the user operation identifier corresponding to the user operation performed next the user operation corresponding to the user operation identifier obtained in the step 1401 and the latest user operation  
10 identifier in the search history table 126.

In the repetitive process, when, in the step 1403, the content of the user operation information corresponding to the user operation identifier is first determined to be an input of an evaluation for the  
15 search result document, the searching profile updating program 117 is started in the step 1406 by using, as an input, a document number and an evaluation described in the user operation information.

Next, in the step 1404, when the content of  
20 the user operation information corresponding to the user operation identifier is determined to be restoration of searching profile, the searching profile restoring program 122 is started in the step 1407 by using an input of the user operation identifier.

25 Subsequently, in the step 1405, when the content of the user operation information corresponding to the user identifier is determined to be erase of user operation, the user operation-erasing program 123 is

started in the step 1408 by using an input of the user operation identifier.

Through the above repetitive process, the searching profile is returned to the state immediately preceding the user operation designated by the user and thereafter, repetition starts with a user operation carried out next the user operation designated by the user. In this manner, all user operations excepting the user operation designated by the user can be reproduced. In other words, only the user operation designated by the user can be erased.

Next, user operation information is prepared on the basis of the content read in the step 1401 and is returned as an output to the search control program 113 that has called the present program. The user operation information now prepared has the content "the user operation of user operation identifier 5 is erased", indicating that the user has erased the user operation. This information is saved in the search history table 126 while being caused to have the correspondence with the searching profile changed in the repetitive process of the steps 1401 and 1402 and the search result based on the searching profile by means of the search history saving program scheduled to be executed after execution of the present program.

The above is the contents of the user operation-erasing program.

Referring now to a PAD of Fig. 15, a

processing procedure in the search session ending program 130 started by the search control program 113 in the step 618 shown in Fig. 5 will be described.

Firstly, in step 2001, the search session  
5 ending program 130 clears the contents of the searching profile saving area 124 and search history table 126.

Next, in step 2002, the files in the searching profile history file group 128 and search result history file group 129 are erased.

10 Through the above steps, the history of the user operations, searching profiles and search results after setting of the source document is cleared to end a series of search sessions.

In this phase, by saving the content of the  
15 searching profile saving area 124 before clearing the content of the searching profile saving area 124 in the step 2001, the saved content can be reutilized for the next and ensuing search sessions.

The processing procedure in the search session  
20 ending program 130 is carried out as described above.

While in the foregoing the processing  
procedures in the individual programs in the present embodiment have been described, a concrete process flow in the relevance feedback in the embodiment will now be  
25 described by making reference to Figs. 16, 17, 18, 19 and 20.

Referring first to Fig. 16, a concrete process flow in the case where when the user initially inputs a

seed document as a search condition, a searching profile is generated to perform a similar document search will be described.

5 Firstly, in step 1701, a searching profile is generated on the basis of a seed document 1700 by means of the searching profile-generating program 114. More specifically, the seed document 1700 inputted by the user is acquired, characteristic strings are extracted from the seed document 1700 and these strings are  
10 assigned with weights so as to be written into the searching profile storing area 124. In an example illustrated in the figure, a seed document "A high-school baseball opens following soccer---" is acquired and characteristic strings such as "soccer", "high-  
15 school", "baseball" and "opens" and their weights in terms of frequencies of appearance in the seed document are written in the searching profile saving area 124.

Next, in step 1702, a similar document search based on the content of the searching profile saving  
20 area 124 is carried out by the similar document searching-program 115 and a search result 1706 is displayed on the display 100.

Subsequently, in step 1703, the content of the searching profile saving area 124 is saved in a file in  
25 the searching profile history saving file group 128 by means of the search history saving program 118. In the illustrated example, the content of the searching profile saving area 124 is saved in a file named

"0001.pfl" in the searching profile saving file group 128.

Then, in step 1704, the search result 1706 is saved in a file named "0001.res" in the search result  
5 history saving file group 129 by means of the search history saving

Thereafter, in step 1705, the search history saving program 118 writes information in the search history table. In the illustrated example, user  
10 operation information to the effect that the user "sets the seed document", the resulting file name "0001.pfl" and the file name "0001.res" saving the search result based on the searching profile are saved in the search history table 126 while being caused to have the  
15 correspondence between them and added with a user operation identifier.

The above is the process flow in the case where the searching profile is generated and the similar document search is conducted when the user initially  
20 inputs the seed document as a search condition in the present embodiment.

Referring now to Fig. 17, a process flow in the case where the user inputs an evaluation for a search result document and the searching profile is  
25 updated on the basis of the evaluation to perform re-retrieval in the embodiment will be described.

In an illustrated example, it is assumed that the user inputs an evaluation for a search result

document obtained as a result of setting the seed document as shown in Fig. 16 to change the searching profile.

5 Firstly, in step 1801, the searching profile is updated by means of the searching profile updating program 117. More specifically, characteristic strings are extracted from a document subjected to evaluation by the user and the content of the searching profile saving area 124 is updated on the basis of the extracted  
10 strings. In the illustrated example, characteristic strings such as "high-school", "baseball" and "Kohshien (the name of ball park)" are extracted from a document of document number 10 evaluated as desirable by the user. Then, the weights of "high-school" and "baseball"  
15 that have already been registered in the searching profile saving area 124 add. "Koshien" that has not been registered till then in the searching profile saving area 124 is newly subjected to additional registration.

20 Next, in step 1802, a similar document search based on the searching profile updated in the step 1801 is conducted by means of the similar document search program 115 and a search result 1806 is displayed on the display 100.

25 Subsequently, in step 1803, the content of the searching profile saving area 124 is saved in a file in the searching profile saving file group 128 by means of the search history saving program 118. In the

illustrated example, the content of the searching profile saving area 124 is saved in a file named "0002.pfl" in the searching profile saving file group 128.

5                   Then, in step 1804, the search result 1806 is saved in a file in the search result history saving file group 129 by means of the search history saving program 118. In the illustrated example, of the search result 1806, orders and document numbers corresponding to the  
10 orders, respectively, are written in a file named "0002.res" in the search result history saving file group 129.

                  Thereafter, in step 1805, information is written into the search history table by means of the  
15 search history saving program 118. In the illustrated example, since the searching profile is updated as a result of the fact that the user evaluates the document of document number 10 as "desirable", user operation information "document 10 is evaluated as desirable", the  
20 file name "0002.pfl" saving a searching 1 profile changed by the evaluation and the file name "0002.res" saving the search result based on the changed searching profile are saved in the search history table 126 while being caused to have the correspondence between them and  
25 added with a user operation identifier.

                  The above is the process flow in the case where the user inputs an evaluation for the search result document and the searching profile is updated on



the basis of the evaluation to conduct re-retrieval in the present embodiment.

Referring now to Figs. 18 and 19, a process flow in the case where after the user carries out  
5 relevance feedback to change the searching profile, the content of a searching profile at a desired time is restored and a search result based on the searching profile is displayed to the user in the present embodiment will be described.

10 In an example shown in Figs. 18 and 19, it is assumed that after performing a similar document search by inputting a seed document and inputting evaluations for search result documents three times, the user makes a request for returning the content of searching profile  
15 to that immediately after inputting of the second evaluation.

Firstly, in step 1901 shown in Fig. 18, the search history table 126 is consulted by means of the search history-displaying program 121 to display a  
20 history of user operations. In the illustrated example, a history of user operations in which after setting the seed document, the user inputs evaluations three times is displayed together with corresponding user operation identifiers.

25 Next, in step 1903, the searching profile restoring program 122 consults the search history table 126 to read a searching profile history saving file corresponding to a user operation identifier inputted by

the user and replace the content of the searching profile saving area 124. In the illustrated example, since an instance is assumed in which the user desires "the searching profile is to be returned to that

5 immediately after document 80 is evaluated as undesirable", the user consults information displayed in the search history displaying picture 1906 to input a corresponding user operation identifier "3". Accordingly, in the present step, a file "0003.pfl" in  
10 searching profile history saving file group 128 corresponding to the user operation identifier "3" in the search history table 126 is read and the content of the searching profile saving area 124 is replaced with the content of the file "0003.pfl".

15 Subsequently, in step 1904 shown in Fig. 19, the searching profile restoring program 122 consults the search history table 126 to read a search result history file corresponding to the user operation identifier inputted by the user and display a search result on the  
20 screen. In the illustrated example, a file "0003.res" in search result history saving file group 129 corresponding to the user operation identifier "3" in the search history table 126 is read and its content, that is, document numbers and title of the corresponding  
25 individual documents are displayed. Through this, a search result based on a searching profile replaced with the content of the file "0003.pfl" can be displayed.

Thereafter, in step 1905, the search history

saving program 118 writes information into the search history table 126. Since, in the illustrated example, the user has returned the searching profile to the state immediately after the user operation of user operation  
5 identifier "3", user operation information "restoration of the state after user operation 3", the file name "0003.pfl" saving a searching profile after restoration by the user operation and the file name "0003.res" saving a search result based on the searching profile  
10 are saved in the search history table 126 while being caused to have the correspondence between them and added with the user operation identifiers.

The above is the process flow in which the user performs the relevance feedback to change the  
15 searching profile and thereafter the content of the searching profile at a desired time is restored to display a search result based on the restored searching profile to the user in the embodiment.

The concrete process flow of the relevance  
20 feedback in the present embodiment is carried out as described above.

Referring now to Fig. 20, there is illustrated an example of a picture displayed by the search history displaying program 121 when the user makes a request for  
25 search history display in the step 614 shown in Fig. 5 in the present embodiment. Assumptively, in the case of Fig. 20, the user inputs an evaluation in accordance with the procedure exemplified in Figs. 16 and 17 and

thereafter, makes the request for search history display.

At that time, a list 1901 of user operations carried out after the user has set an initial seed document is displayed. When the user operation is an evaluation for a search result document as shown in this figure, a document number of the document subjected to evaluation may be consulted to display an attribute such as title of the document.

The user can input a user operation identifier to a text box 1902 and depress any one of button 1903 for "return the state to the immediately succeeding searching profile", button 1904 for "erase the user operation" and button 1905 for "display the immediately succeeding search result".

When the button 1903 is depressed, the searching profile restoring program 122 is started to restore a searching profile immediately succeeding a searching profile changed by a user operation corresponding to a user operation identifier inputted to the text box 1902.

With the button 1904 depressed, the user operation-erasing program 123 is started to change the searching profile to the state when the user does not execute a user operation corresponding to a user operation identifier inputted to the text box 1902.

With the button 1905 depressed, the search result history-displaying program 120 is started to

display a search result based on a searching profile immediately succeeding a searching profile changed by a user operation corresponding to a user operation identifier inputted to the text box 1902.

5           If a "recommencement" button is provided and this button is depressed, it may be considered that the latest user operation identifier is designated and a searching profile preceding execution of the immediately preceding user operation may be restored.

10           The above is the example of a picture displayed by the search history-displaying program 121 in the present embodiment.

          As described above, according to the present embodiment, a desired evaluation designated by the user  
15   can be recommenced to return the retrieval profile to that at a desired time. Accordingly, a search result document based on the returned searching profile can be presented and recommencement of an evaluation for the search result document can be permitted. In other  
20   words, improvements in accuracy of retrieval based on the trial and error relevance feedback can be realized.

          In the present embodiment, document retrieval is carried out for the precedently registered document database but by providing search history tables 126 by  
25   the number of a plurality of users, the present invention can also be applied to a text dipping system in which a document coinciding with a searching profile registered by each user can be delivered to each user.

In other words, the present invention can be applied to a system in which electronic documents acquired from an information source such as communication company or newspaper company by using, for example, an electronic  
5 mail or news collection robot are retrieved by means of searching profiles of a plurality of precedently registered users and documents satisfying a predetermined condition are distributed to the individual users.

Further, dates/times for inputting evaluations  
10 may be saved in the search history table 126. In this case, a searching profile of a date/time desirably designated by the user can be restored.

Further, in the search result history displaying program 120, the saved contents of files in  
15 the search result history file group 129 may all be merged so as to display all search results acquired in the past in a list form. In this case, documents evaluated by the user through collation with the content of the search history table 126 may be displayed  
20 together with evaluation information and user operation identifiers. When the user designates a user operation identifier by consulting this display, a desired evaluation can be recommenced or return to a searching profile at a desired time can be ensured as explained in  
25 connection with Fig. 18.

Further, in the present embodiment, the user inputs an evaluation for one search result document but evaluations for more than one search result documents

can be inputted at a time and changing to a searching profile reflecting these evaluations may be effected. In this case, the same user operation identifier can be assigned to the plurality of evaluations inputted at a  
5 time and saved in the search history table.

In the present embodiment, a seed document is initially set and a document having the content similar to that of the seed document is retrieved but a scheme for retrieval of the whole passage in which a keyword is  
10 initially set may be employed. In that case, in place of the step 702 shown in Fig. 6, a process may be involved in which the inputted keyword is applied with a weight through a predetermined method and is additionally saved in the searching profile saving area 124.

15 According to the embodiment set forth so far, in a document retrieval method having relevance feedback, a document evaluated by the user and a history of evaluations are managed by making the correspondence of them with searching profiles at times for evaluations  
20 and therefore recommencement of relevance feedback at a desired time can be ensured.

As described above, according to the present invention, recommencement of evaluation for a desired search result document can be permitted to thereby  
25 improve the retrieval accuracy.